

Jörg Porsiel (Wolfsburg)

Maschinelle Übersetzung bei der Volkswagen AG



Editors:

Viktorija Bilić

Anja Holderbaum

Anne Kimmes

Joachim Kornelius

John Stewart

Christoph Stoll

Publisher:

Wissenschaftlicher Verlag Trier

Jörg Porsiel (Wolfsburg)

Maschinelle Übersetzung bei der Volkswagen AG

Abstract:

Der Artikel stellt den Einsatz maschineller Übersetzung in einem weltweit tätigen Unternehmen der Automobilindustrie dar. Wichtige Faktoren in Bezug auf den effizienten Einsatz von und verantwortungsvollen Umgang mit maschineller Übersetzung sind kontinuierliche bewussteinbildende Maßnahmen für deren Nutzer. Ziele sind dabei u. a. die Darstellung des individuellen Nutzens sowie die dauerhafte Verbesserung der Übersetzungsqualität auf der einen Seite. Auf der anderen Seite stehen grundlegende Datenschutzerwägungen und die Bewusstmachung der meist völlig verkannten Risiken maschineller Übersetzung in Verbindung mit Angeboten aus dem Internet.

This article describes the use of machine-translation in a global automotive company. Efficient and responsible use of machine translation requires ongoing supportive activities that raise the awareness of these technology users. The objectives of these activities include highlighting the benefits for the individual translator, as well as consistently improving the quality of translation solutions. Further emphasis is placed on the fundamental importance of data security and the often underestimated risks of machine translation in conjunction with online solutions.

Keywords:

maschinelle Übersetzung, Terminologiemanagement, kontrollierte Sprache, Datensicherheit, Industriespionage, Bewussteinbildung, zielgruppengerechtes Schreiben, übersetzungsgerechtes Schreiben, Cloud-Computing, Crowdsourcing, management of expectations

machine translation, terminology management, controlled language, data security, industrial espionage, raising awareness, writing for a target group, writing for translation, cloud computing, crowd sourcing, management of expectations

Inhalt:

1	Ausgangslage und Rahmenbedingungen bei Volkswagen.....	2
2	Maschinelle Übersetzung bei Volkswagen.....	2
2.1	Anwendungsspektrum und „management of expectations“	4
2.2	Was maschinelle Übersetzung kann und was nicht.....	7
2.3	Warum maschinelle Übersetzung „so schlecht“ ist	8
3	Maschinelle Übersetzung und Datensicherheit.....	9
3.1	„Kostenlose“ maschinelle Übersetzung im Internet.....	10

3.2	Google Translate, Microsoft Bing, Yahoo! Babel Fish & Co.	11
3.3	Industriespione, Hacker, Produktpiraten etc.....	12
3.4	Maschinelle Übersetzung als Tatwerkzeug.....	13
4	Ausblick	15
5	Bibliografie	16

1 Ausgangslage und Rahmenbedingungen bei Volkswagen

Der Volkswagen Konzern, der 2011 weltweit rund 500.000 Mitarbeiter zählte, davon etwa 100.000 in Deutschland, bietet seine Fahrzeuge in 153 Ländern an. An 62 Produktionsstandorten werden Fahrzeuge der Marken Volkswagen, Audi, Seat, Škoda, Bentley, Lamborghini, Bugatti, Porsche, Ducati, Volkswagen Nutzfahrzeuge, MAN und Scania gefertigt. 2011 lieferte Europas größter Autobauer 8,36 Millionen Fahrzeuge aus, wobei sich der Umsatz auf 159,3 Milliarden Euro belief.

Für die Fahrzeuge werden verschiedene Arten von „Literatur“ benötigt, zum einen für die Fahrzeughalter, zum anderen für das Werkstattpersonal. Für die erste Gruppe ist das „Bordbuch“ bestimmt, die Gebrauchsanleitung für das jeweilige Fahrzeug. Für die zweite Gruppe sind es z. B. so genannte „Reparaturleitfäden“ für Fahrzeugwartung und -reparatur. Diese unterschiedlichen Textsorten, die zudem in unterschiedlichen Dateiformaten und Darreichungsformen erstellt und verteilt werden (on- und offline), müssen in bis zu 40 Sprachen übersetzt werden, wobei fast durchgängig Deutsch die Ausgangssprache ist. Das Volkswagen-eigene *Fremdsprachenmanagement* in Wolfsburg wickelte 2010 ca. 30.000 Übersetzungsaufträge sowie ca. 1.200 Dolmetschereinsätze ab. Das Textvolumen umfasste dabei ca. 340 Millionen Wörter. Auf DIN-A4 gedruckt und an der Schmalseite aneinander gereiht, würden diese Seiten ein Mal den Äquator umspannen.

2 Maschinelle Übersetzung bei Volkswagen

Seit 2002 stellt das Volkswagen Fremdsprachenmanagement im konzerneigen und weltweit für Mitarbeiter zugänglichen Intranet maschinelle Übersetzung (MÜ) zum dienstlichen Gebrauch zur Verfügung. Die Anwendung ist Bestandteil des „Volkswagen Sprachenportals“,

das die erste Anlaufstelle und Präsentationsplattform für die Abteilung „Fremdsprachenmanagement“ ist. Über diese Website haben die Mitarbeiter u. a. auch Zugriff auf eine unternehmenseigene, von Terminologen gepflegte Terminologiedatenbank mit etwa 20.000 VW-spezifischen Begriffen in 10 Sprachen.

Abbildung 1: Startseite des Volkswagen Sprachportals

Nach einer Testphase hat sich Volkswagen für ein so genanntes „regelbasiertes“ System der deutschen Firma Lucy Software and Services GmbH entschieden, das mittlerweile in den Sprachen Deutsch, Englisch, Französisch, Russisch und Spanisch zur Verfügung steht. In den letzten Jahren verstärkt auf den Markt drängende statistische oder Hybrid-Systeme, waren 2002, zum Zeitpunkt der Einführung bei Volkswagen, noch nicht marktreif.

Grob vereinfacht, lassen sich zwei große „Schulen“ oder Ansätze maschineller Übersetzung unterscheiden: Der „klassische“, regelbasierte Ansatz (RBMT = rule-based machine translation) versucht vor der eigentlichen Verarbeitung/Übersetzung die linguistischen

Informationen des Ausgangstextes, d. h. dessen syntaktische, morphologische und semantischen Strukturen mittels für jede Sprache und Sprachrichtung speziell entwickelte Algorithmen zu analysieren und so Texte ohne Informationsverlust in die Zielsprache zu übertragen. „Klassisch“ ist dieser MÜ-Ansatz insofern, als er vor rund 60 Jahren entwickelt wurde. Vor etwa 15 Jahren kam ein neuer Ansatz hinzu, die „statistische maschinelle Übersetzung“ (SMT = statistical machine translation). Sie versucht, Text auf Grundlage statistischer Wahrscheinlichkeiten zu übersetzen. Ein in jüngster Zeit entstandene Kombination aus beiden Ansätzen ist die so genannte „Hybride maschinelle Übersetzung“ (hybrid machine translation), bei der versucht wird, die Vorteile der anderen beiden Ansätze zu kombinieren und so bessere Ergebnisse zu erzielen.

2.1 Anwendungsspektrum und „management of expectations“

Ausschlaggebend für die Bereitstellung maschineller Übersetzung im VW Intranet waren mehrere Faktoren: Zum einen die wachsende Notwendigkeit, Texterstellungs- und Übersetzungsprozesse zu beschleunigen und gleichzeitig qualitativ zu verbessern. Dazu soll den Mitarbeitern bewusst gemacht werden, welche (betriebswirtschaftliche) Bedeutung eine einheitliche, eindeutige Unternehmenssprache hat. Darüber hinaus soll umfassendes Terminologiemanagement in der technischen Dokumentation weiter vorangetrieben werden. Diese Ziele sollten durch die Bereitstellung einer möglichst intuitiv bedienbaren Anwendung zur Bewältigung des kontinuierlich steigenden E-Mail-Aufkommens sowohl inner- als auch außerhalb des Konzerns erreicht werden. Darüber hinaus musste ebenfalls der starke Anstieg des Textvolumens der technischen Dokumentation aufgrund der Ausweitung der Fahrzeugmodellpalette sowie der Erschließung neuer Märkte bewältigt werden.

Weltweit werden täglich mehrere Milliarden E-Mails versandt. Im Durchschnitt erhält ein Mitarbeiter eines in den Fortune 500 gelisteten Unternehmens, wozu auch die Volkswagen AG gehört, pro Tag 80 E-Mails. Schätzungen zufolge verlieren 70 % der Mitarbeiter eines Großunternehmens in Deutschland täglich 60 Minuten ihrer Arbeitszeit allein durch die zeitaufwändige Bearbeitung „unverständlicher“ E-Mails (dazu gehören in diesem Fall auch

fremdsprachige). Zudem führt die steigende Zahl der E-Mails nicht zwangsläufig zu einer Verbesserung der Kommunikation bzw. des Informationsaustausches.

Ineffizienter Informationsaustausch durch unverständliche E-Mails kann auch darin bestehen, dass fremdsprachige E-Mails aufgrund unzureichender Sprachkenntnisse nicht bzw. missverstanden werden. Auch können derartige E-Mails dazu führen, dass sie wegen der Wartezeit für eine notwendige Humanübersetzung (zu) spät beantwortet oder aber falsche Maßnahmen aus ihnen abgeleitet werden. Hier sorgt maschinelle Übersetzung für erhebliche Prozessbeschleunigung: In Sekundenschnelle können Textmengen von einer Zeile bis hin zu mehreren Hundert Seiten übersetzt werden – wobei die Übersetzungsqualität zunächst von untergeordneter Bedeutung ist. Zeitgewinn ist das vorrangige Ziel beim Einsatz maschineller Übersetzung!

Volkswagen setzt maschinelle Übersetzung seit 2002 konzernweit zur beschleunigten Abarbeitung von Routineaufgaben ein. Dabei geht zunächst um „gist translations“ oder „indicative translations“, also Roh- oder Informationsübersetzungen. Obwohl deren Verständlichkeit in Abhängigkeit von der Qualität des Ausgangstextes (GIGO-Prinzip, s. dazu unten „Warum maschinelle Übersetzung ‚so schlecht‘ ist“) zum Teil stark schwanken kann, reicht den Anwendern das Übersetzungsergebnis in der Regel aus, um innerhalb von Sekunden und in Verbindung mit dem eigenen Vor- und Fachwissen eine Entscheidung treffen zu können, wie mit der erhaltenen Information weiter zu verfahren ist.

Aus der Konstellation von Informationsbedürfnis, -beschaffung und -verarbeitung – in Verbindung mit dem enormen Zeitgewinn durch MÜ – resultiert der primäre Nutzen für das Unternehmen. Insbesondere bei einem weltweit über alle Zeitzonen hinweg operierenden Unternehmen wie Volkswagen ist der Faktor *Zeit* entscheidend. Der Einsatz von MÜ bei Routineaufgaben, wie dem Lesen, Beantworten oder Weiterleiten von E-Mails o. Ä., führt zu einer bisher nicht gekannten Beschleunigung fremdsprachiger Kommunikations- und Entscheidungsfindungsprozesse, verbunden gleichzeitig mit einer erheblichen und nachhaltigen Senkung von Bearbeitungs- und Übersetzungskosten.

Von zunehmender Bedeutung für ein weltweit agierendes Unternehmen wie die Volkswagen AG ist allerdings seit einiger Zeit auch der Aspekt Datensicherheit, der sich aus der Verwendung von maschineller Übersetzung in einem gesicherten Intranet im Gegensatz zu einer (unbekannten) MÜ-Anwendungen irgendwo in den unbekanntenen Weiten des Internets ableitet.

Maschinelle Übersetzung

Online-Übersetzung

Fremdsprachenmanagement

Deutsche Rechtschreibung

Links

Redaktion

Umsetzung

People can't share knowledge if they don't speak a common language.
(Davenport/Prusak: Working Knowledge, Harvard Business School Press, 1998)

Man kann Wissen nur mit jemandem teilen, wenn man dieselbe Sprache spricht.

Und so funktioniert's:

1. Wählen Sie eine Übersetzungsrichtung.
2. Kopieren Sie Ihren Text in das "Texteingabefeld".
3. Klicken Sie auf "Übersetzen".
4. Das Ergebnis erscheint in einem neuen Fenster.

Grundsätzliches: [Möglichkeiten und Grenzen maschineller Übersetzung](#)

Bitte wählen Sie eine Übersetzungsrichtung aus:

<input checked="" type="radio"/> Deutsch → Englisch	<input type="radio"/> Englisch → Deutsch
<input type="radio"/> Deutsch → Russisch	<input type="radio"/> Russisch → Deutsch
<input type="radio"/> Deutsch → Französisch	<input type="radio"/> Französisch → Deutsch
<input type="radio"/> Deutsch → Spanisch	<input type="radio"/> Spanisch → Deutsch

Texteingabefeld

Bitte beachten Sie die neue deutsche Rechtschreibung, korrekte Groß- und Kleinschreibung und Zeichensetzung.

Haftungsausschluss
Übersetzen

Abbildung 2: Startseite der maschinellen Übersetzung

Maschinelle Übersetzung wird seit ihrer Einführung im Volkswagen Konzern zur dienstlichen Nutzung bereitgestellt. Werktätlich, also von Montag bis Freitag, werden gegenwärtig ca. 12.000 Übersetzungsaufträge in den fünf Sprachen verarbeitet, wobei Englisch als Ziel- und Ausgangssprache mit weitem Abstand vor Spanisch rangiert. In Standardseiten umgerechnet entsprechen die 12.000 Übersetzungen pro Arbeitstag etwa 2.300 DIN-A4-Seiten.

2.2 Was maschinelle Übersetzung kann und was nicht

Was ein Mensch in seiner Muttersprache in der Regel problemlos schafft, ist für maschinelle Übersetzung „tödlich“:

Gmäß eneir Sutide eneir elgnihcesn Uvinisterät ist es nchit witihcg, in wlecehr Rneflogheie die Bstachuebn in eneim Wrot snid, das ezniige was wcthiig ist, ist, dass der estre und der leztte Bstabchue an der ritihcegn Pstioion snid. Der Rset knan ein ttoaelr Bsinöldn sien, tedztorm knan man ihn onhe Pemoblre lseen. Das ist so, wiel wir nciht jeedn Bstachuebn enzelin leesn, snderon das Wrot als gseatems. ([Wikipedia](#))

Der Mehrheit der Anwender ist nicht bewusst, was MÜ zu leisten vermag und was nicht und warum. So ist Bewusstseinsbildung auf Seiten der Nutzer in Bezug auf Möglichkeiten und Grenzen maschineller Übersetzung wichtig, denn nichts ist für die Akzeptanz maschineller Übersetzung schädlicher als schlechte Ergebnisse durch frustrierte Anwender. Es muss also deutlich gemacht werden, dass die Qualität der Ausgangstexte – natürlich immer in Verbindung mit entsprechend hochwertiger Terminologie – für die Übersetzungsqualität von entscheidender Bedeutung ist. Darüber hinaus muss aber ebenfalls deutlich dargestellt werden, was MÜ nicht kann, nämlich: zwischen den Zeilen lesen, ahnen, interpretieren, vermuten, assoziieren, Nicht-Übersetzbares paraphrasieren und vor allem fehlertolerant sein.

Beispiel für einen „guten“ Ausgangstext:

In Europa verkaufte Volkswagen Pkw bis April 2012 589.800 Fahrzeuge (567.000 im Vorjahr) und erreichte damit ein Plus von 4 %. Die Marke konnte dabei vor allem in Zentraleuropa und Osteuropa wachsen, wo die Auslieferungen deutlich um 55,7 % auf 84.000 Fahrzeuge (53.900) anstiegen. In Russland, dem größten Einzelmarkt der Region, verdoppelten sich die Verkaufszahlen auf 50.500 Einheiten (24.900; +103,0 %). In Westeuropa, ohne Deutschland, gingen die Auslieferungen hingegen aufgrund der schwierigen Situation des Gesamtmarktes um 5,1 % auf 303.400 Einheiten zurück (319.800). In Deutschland hingegen steigerten sich die Auslieferungen um 4,7 % auf 202.400 Fahrzeuge (193.300).

Volkswagen passenger cars sold in Europe to April 2012 589,800 vehicles (567,000 in the previous year) and a plus of 4 % reached in this way. The [brand|mark|stamp] could increase in this case particularly into Central Europe and Eastern Europe, where the [deliveries|extraditions|handing over] went up clearly around 55.7 % onto 84,000 vehicles (53,900). In Russia, the biggest single market of the [region|area], the sales figures doubled on 50,500 units (24,900; +103,0 %). On the other hand in Western Europe, without Germany, the [deliveries|extraditions|handing over] decreased about 5.1 % due to the difficult situation of the total market on 303,400 units (319,800). In Germany on the other hand increased the [deliveries|extraditions|handing over] around 4.7 % onto 202,400 vehicles (193,300).

Abbildung 3: Übersetzungsergebnis bei guter Ausgangstextqualität

Das Übersetzungsergebnis ist sicherlich noch nicht optimal, dennoch erfüllt es den Zweck einer Roh- oder Informationsübersetzung. Für einen geschulten Post-Editor wäre sie einfach und schnell zu bearbeiten, sodass sie – sofern gewünscht – anschließend publiziert werden könnte. Für den „Hausgebrauch“ als Informationsübersetzung dürfte sie aber in diesem Zustand bereits vollkommen ausreichen.

2.3 Warum maschinelle Übersetzung „so schlecht“ ist

Gelegentlich beschwerten sich Anwender, dass die Ergebnisse der maschinellen Übersetzung „so schlecht“ seien. Besonders zwei „Beweise“ führen sie dabei oft an: 1. Beweis: Die Übersetzung einzelner Begriffe „funktioniert nicht“. „Ich habe ‚xy‘ eingegeben, aber keine Übersetzung erhalten.“ oder „... aber eine völlig falsche Übersetzung erhalten.“; und 2. Beweis: Einige Nutzer sind der Meinung, das ideale Messmittel für Übersetzungsqualität gefunden zu haben: nämlich die Rückübersetzung. Wie bei der Gegenprobe in der Mathematik wird der gerade übersetzte Text in die Ausgangssprache zurückübersetzt. Ist das Ergebnis dann nicht vollkommen oder doch wenigstens zu mindestens 90 % identisch mit dem Ursprungstext, handelt es sich nach Ansicht dieser Nutzer zweifelsfrei um eine „schlechte Übersetzung“.

Es bleibt festzustellen, dass eine der Hauptfehlerquellen maschineller Übersetzung der Anwender selbst ist. Häufig werden unvollständige Sätze eingegeben, solche mit falscher oder gar keiner Interpunktion, fehlerhafter Rechtschreibung, (in Deutsch) absichtlicher

genereller Kleinschreibung oder mit willkürlichen Abkürzungen. Sehr häufig kommen Texte vor, die größtenteils aus einem für einen Computer unverarbeitbaren Sprachgemisch, wie z. B. Denglisch, bestehen – Textqualitäten also, bei denen selbst ein Übersetzer aus Fleisch und Blut erhebliche Schwierigkeiten hätte, überhaupt zu verstehen, was gemeint sein könnte.

Beispiel eines „schlechten“ Ausgangstextes (1:1 aus einer Log-Datei entnommen):

habe u.g. HWMotor 6J0.955.711 hinsichtl. ENBaulage an drei Bsp. aktueller VW-Neuprojekte
geengeprüft. demnach ist dieser HWMot. für uns zur Anfrage

[u.g] has. [HWMotor 6J0.955.711] with regard to [ENBaulage] at three [Bsp].
current VW new projects' tested. accordingly this [HWMot] is. for us for the inquiry

Abbildung 4: Übersetzungsergebnis bei schlechter Ausgangstextqualität

Jeder Profi-Übersetzer weiß, dass die Qualität des Ausgangstextes von entscheidender Bedeutung für die Qualität des Endproduktes ist. Dies gilt für MÜ jedoch in besonderem und viel gravierenderem Maße, denn bei ihr wird nicht (oder sozusagen „falsch“) verarbeitet, was unbekannt ist – und sei es auch nur, weil ein Wort falsch geschrieben wurde. Hier schlägt das „garbage-in-garbage-out“-Prinzip (GIGO) voll durch. Der Einsatz von MÜ zur Erzeugung qualitativ hochwertiger maschineller Übersetzungen (HQMT – High Quality Machine Translation) für vordefinierte Textsorten unter Verwendung unternehmenseigener Terminologie in Verbindung mit Translation Memories und Prä- und Posteditingsprozess erfordert ein hohes Maß an Planung und Qualifikation seitens der den Prozess betreuenden Mitarbeiter.

3 Maschinelle Übersetzung und Datensicherheit

Obwohl viele der Meinung sind, maschinelle Übersetzung sei die Lösung aller Übersetzungsprobleme (bei gleichzeitiger Reduzierung der eigenen Übersetzungskosten auf Null), ist es die verantwortungsvolle Aufgabe eines Betreibers – noch dazu, wenn die Anwendung speziell konfiguriert im weltweit zugänglichen Intranet des Unternehmens bereit gestellt wird –, den Nutzern Möglichkeiten und Grenzen von MÜ deutlich vor Augen

zu führen. MÜ kann durchaus bei der Bewältigung bestimmter Aufgaben eine große Hilfe sein, so z. B. für bestimmte Textsorten und Zielgruppen. Es gibt Einsatzszenarien, in denen MÜ gut geeignet ist und andere, in denen sie auf unabsehbare Zeit ungeeignet sein wird. So oder so bedarf es aber veränderter Übersetzungsansätze und -prozesse: Der Übersetzer(job) der Zukunft wird je nach Zielsetzung ein anderer sein, als der, den die meisten heute kennen.

Die geradezu explodierenden Nutzungszahlen von Google Translate, Microsoft Bing Translator und anderen Anbietern demonstrieren eindrucksvoll, dass Bedarf vorhanden, dieser zudem unvorstellbar groß ist und noch weiter wachsen wird. Neben neuen Funktionen und Sprachpaaren erscheinen auch immer mehr neue Anbieter mit z. T. exotischen Sprachen auf der Bildfläche, die ihre Dienste gelegentlich auch im Internet anbieten. Aus Nutzersicht ist die Situation insofern ideal, als im Internet alles kostenlos ist (jedenfalls scheinbar, siehe dazu weiter unten). Den meisten Nutzern maschineller Übersetzung im Internet fehlt jegliches Bewusstsein in Bezug auf Gefahren oder Probleme bei der Verwendung derartiger Angebote. Selbst der überwiegenden Mehrheit der IT-Sicherheitsbeauftragten großer Unternehmen fehlt dieses Bewusstsein, wie aus der unten erwähnten aktuellen Studie (s. u. „Industriespione, Hacker, Produktpiraten etc.“) deutlich hervorgeht.

3.1 „Kostenlose“ maschinelle Übersetzung im Internet

Eine von der breiten (Internet-)Öffentlichkeit vollkommen verkannte oder zumindest doch stark unterschätzte Gefahr ist jene, die von (scheinbar) kostenlosen Angeboten maschineller Online-Übersetzungen ausgeht. In Zeiten von Identitätsdiebstahl, Harvesting, Phishing, Skimming, Snarfing, Social Engineering etc. sowie Cyberattacken von so genannter Malware à la Stuxnet, Duqu oder Flame und dergleichen, muss sich der Internetnutzer Folgendes vergegenwärtigen: Anbieter kostenloser MÜ-Dienstleistungen gibt es im Internet in ständig wachsender Zahl. Neben den „Großen“ wie Google, Bing & Co. gibt es eine schnell wachsende Zahl von mehr oder weniger dubiosen Anbietern, die auch wenig verbreitete Sprachen anbieten.

Wenn man sich bewusst macht, was für ein Entwicklungs- sowie kontinuierlicher Pflege- und Optimierungsaufwand hinter einem solchen System, sei es nun regelbasiert oder statistisch, steckt, muss man sich im Gegenzug fragen, aus welchem Grund der Anbieter diese Dienstleistung kostenlos zur Verfügung stellt und was sein – betriebswirtschaftlicher oder sonstiger – Nutzen ist. Es wäre naiv zu glauben, dass Altruismus der Motivator ist.

3.2 Google Translate, Microsoft Bing, Yahoo! Babel Fish & Co.

Anfang 2012 änderte Google die Nutzungsbedingungen für seine Online-Anwendungen wie folgt:

When you upload or otherwise submit content to our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content. The rights you grant in this license are for the limited purpose of operating, promoting, and improving our Services, and to develop new ones. This license continues even if you stop using our Services [...].
([Google Terms of Service](#)).

Das wiederum ist nichts anderes als ein Freibrief für Google mit den von den Nutzern seiner Angebote zur Verfügung gestellten Daten zu machen, was auch immer Google damit zu tun gedenkt – ungeachtet übrigens der Frage, ob der Nutzer tatsächlich das (Urheber-)Recht an den zur maschinellen Übersetzung eingestellten Texten besitzt. Aber das ist ein ganz anderes rechtliches Problem. Buchstäblich jedes Zeichen und jedes Wort, das jemand bei Google Translate übersetzen lässt wird – inklusive natürlich der von Google selbst erzeugten Übersetzung – ist Eigentum von Google und kann somit nach Gutdünken und Belieben für aktuelle – und zukünftige – Unternehmensanwendungen weiter verarbeitet werden.

Im schlimmsten Fall „bezahlt“ der unwissende oder unaufmerksame Anwender also diesen vermeintlich „kostenlosen“ Service mit (seinen eigenen) persönlichen und/oder wirtschaftlichen Daten, indem er sie Google zur unentgeltlichen und dauerhaften Nutzung überlässt und dafür auch noch implizit seine Einwilligung (durch die Nutzung des bereitgestellten Online-Angebotes) gegeben hat.

3.3 Industriespione, Hacker, Produktpiraten etc.

Industriespionage ist schon lange weltweite Realität. Nach Untersuchungen der Sicherheitsberatungsfirma Corporate Trust sowie des Verbandes Deutscher Maschinen- und Anlagenbau (VDMA) vom Frühjahr 2012 entstehen der deutschen Wirtschaft durch Angriffe von Hackern sowie durch – wie auch immer gearteten – Geheimnisverrat jährlich Schäden von 4,2 Mrd. Euro. In der vorausgegangenen Studie aus dem Jahre 2007 waren es noch 2,8 Mrd. Euro Schaden. Der VDMA geht davon aus, dass allein durch Produktpiraterie ein Schaden von 8 Mrd. Euro im Jahre 2011 entstanden ist, 25 % mehr als nur ein Jahr zuvor. In Arbeitsplätze „umgerechnet“ wären das ca. 37.000 (heise.de).

Nach Aussage des deutschen Bundesamtes für Sicherheit in der Informationstechnik (BSI) und des VDMA gewinnt das Internet dabei als „bequemes“ und vergleichsweise „komfortables“ Tatwerkzeug für ausländische Nachrichtendienste, Konkurrenzunternehmen, die organisierte Kriminalität und illoyale eigene Mitarbeiter (letztere immerhin 47,8 % aller Schadensfälle) immer größere Bedeutung. So stellte das deutsche Bundeskriminalamt (BKA) in seiner aktuellen Kriminalstatistik für das Jahr 2010 fest, dass zwar die (bekannt gewordene) Zahl von Straftaten um zwei Prozent sank, die von „Ausspähen [und] Abfangen von Daten“ im selben Zeitraum aber um 32,2 % stieg, wobei die Verwendung des Internets als Tatwerkzeug gleichzeitig um 190 % nach oben schoss ([Corporate Trust: 5](#)).

Konzertierte, groß angelegte und professionell gesteuerte Hackerangriffe sind mittlerweile eine alltägliche Erscheinung. Cyberangriffe der jüngsten Zeit auf Sony, Google, die NATO oder den Internationalen Währungsfonds (IWF) werden längst nicht mehr von (minderjährigen) Einzeltätern, Nerds oder Geeks begangen, sondern von IT-Spionage-Profis, die z. T. von Regierungen gesteuert werden, von denen, laut VDMA, die chinesische an erster Stelle steht, gefolgt von den GUS-Staaten ([Corporate Trust: 17](#)).

Laut o. g. Studie sind sich viele deutsche – v. a. mittelständische – Unternehmen weder dieser Bedrohung als solcher noch ihres Ausmaßes auch nur annähernd bewusst. Zwar hat jedes zweite deutsche Unternehmen Erfahrungen mit Hackerangriffen und Geheimnisverrat gemacht, aber trotzdem verwendet nur jedes sechste Unternehmen

Datenverschlüsselungstechnologien. Datendiebstahl, unkontrollierte Informationsweitergabe sowie (u. U. sogar ungewollter) Geheimnisverrat durch eigene Mitarbeiter waren 2011 in 47,8 % aller Fälle die Ursache. In Verbindung mit Informationen, die Mitarbeiter unbemerkt durch Social Engineering weitergeben, wächst dieser Wert schließlich auf 70,5 %. Die Top-3-Ziele solcher Angriffe sind laut Studie mit 18,3 % der Vertrieb, gefolgt von 16 % bei Forschung & Entwicklung sowie Mergers & Aquisitions mit 14,2 %.

3.4 Maschinelle Übersetzung als Tatwerkzeug

Der Leidensdruck eines Arbeitnehmers, einen Text schnellstmöglich übersetzt zu bekommen, um zumindest den Informationsgehalt erschließen können, muss nur groß genug sein, damit er sich im Internet auf die Suche nach maschineller Übersetzung macht. Hat er dann eine „Maschine“ gefunden (virtuelle Schleichwege zur Umgehung von URL-Sperren gibt es viele), wird der Text verschickt. Man erhält eine Übersetzung und fertig! Tatsächlich? Nutzer maschineller Übersetzungsangebote im Internet („Internetübersetzung“), sollten jedoch vorher die Antworten auf folgende Fragen parat haben: Wer ist der MÜ-Anbieter? In welchem Land befindet sich der Anbieter? Wo stehen dessen Server? Wer hat Zugriff darauf? Warum bietet er diese Dienstleistung (kostenlos) an? Was geschieht eigentlich technisch mit meinen Daten? Die Wahrscheinlichkeit, alle Fragen beantworten zu können, dürfte kleiner gleich Null sein. Demzufolge sollte sich jeder vorher gut überlegen, ob dieser Dienst überhaupt in Anspruch genommen werden soll und wenn ja für welche Art von Texten/Inhalten.

Angesichts des bisher zum Thema (mangelhafte) Datensicherheit in und unkontrollierter Datenabfluss aus einem Unternehmen Gesagten, ließe sich beispielhaft folgendes Szenarium konstruieren:

Ein Mitarbeiter eines international tätigen Unternehmens erhält eine E-Mail mit einem verschlüsselten Dateianhang. Warum ist der Anhang eigentlich verschlüsselt? Enthält er etwa vertrauliche oder gar geheime Informationen, die nicht für jedermann bestimmt sind? Der Empfänger entschlüsselt die Datei und stellt fest, dass er sie nicht verstehen kann, da sie

in einer Sprache geschrieben ist, die er nicht oder nur unzureichend beherrscht. Aber: Er kennt den Absender und weiß, dass es um ein wichtiges Projekt geht (sonst würde er schließlich keine verschlüsselte E-Mail erhalten). Der Mitarbeiter hat keine Zeit oder keine Mittel, den Text selbst zu übersetzen und sucht sich stattdessen einen MÜ-Anbieter im Internet. Nachdem er diesen gefunden hat, lässt er den geheimen Text entschlüsselt (Natürlich ENT-schlüsselt! Sonst könnte er ja nicht verarbeitet werden!) online über eine ungeschützte Datenleitung übersetzen und erhält Sekunden später die so dringend benötigte Übersetzung zurück. „Das war’s!“, denkt er höchstwahrscheinlich. War’s das tatsächlich?

Die Mehrheit der Nutzer von Internetübersetzungen verschwendet keinen Gedanken an das, was da gerade technisch passiert ist bzw. sein könnte. Viele stellen sich den Prozess vielleicht so vor wie beim Schreiben und Speichern einer ganz normalen Datei: Wenn ich zwei Dateien mit demselben Namen speichern will, überschreibt die aktuellere die ältere und Letztere ist damit „verschwunden“. Demzufolge gehen wahrscheinlich viele einfach davon aus, dass die Datei oder der Text, den sie als Übersetzung erhalten haben, die Originaldatei oder die Originaldaten sind, die sozusagen nur in einer anderen „Erscheinungsform“ von der Maschine zurückgeschickt wurden. Also, denkt der Nutzer, habe ich keine Datenverluste und auch keine (Daten-)Spuren hinterlassen. Im günstigsten Fall denkt er vielleicht, dass der Ausgangstext während der Verarbeitung sozusagen im Internet-Nirvana „verdunstet“ ist.

Dem ist aber nicht so! Tatsächlich ist auf dem Server des Anbieters der Internetübersetzung wahrscheinlich gerade Folgendes geschehen: Der zu übersetzende Ausgangstext wurde zusammen mit der Uhrzeit (time stamp) gespeichert, darüber hinaus wurde ebenfalls die IP-Adresse des Absenders und/oder die URL des Unternehmens gespeichert. Die Internetübersetzung wurde ebenfalls zusammen mit Uhrzeit und IP-Adresse des Empfängers gespeichert. Zusätzlich ist denkbar, dass mindestens eine weitere Datei erzeugt und gespeichert wurde, die nämlich diejenigen Begriffe enthält, die der maschinellen Übersetzung (bisher) unbekannt waren. All diese Daten lassen sich problemlos und beliebig filtern und miteinander verknüpfen.

Nun ist es – je nach den Absichten und der kriminellen Energie des Anbieters – denkbar, dass die so gesammelten Daten mit speziellen Programmen ausgewertet und nach Belieben rekombiniert werden, z. B. nach bestimmten Suchworten, IP-Ranges oder URLs. Je nach Inhalt der Ausgangstexte lassen sich in Kombination mit der IP-Adresse des Absenders/Empfängers und mit relativ geringem IT-Aufwand Themen- oder gar Personenprofile erstellen und analysieren. Personenprofile deshalb, weil die Ausgangstexte u. U. personenbezogene oder -beziehbare Daten wie z. B. Namen, Anschriften, E-Mail-Adressen, Personalnummern, Bankverbindungen, Vertragsdaten, Fristen, Zahlungs- und Währungsangaben etc. enthalten.

4 Ausblick

Einsatz und Nutzung der maschinellen Übersetzung im Volkswagen Konzern werden in den kommenden Jahren kontinuierlich ausgebaut. So werden weitere Sprachen wie z. B. Chinesisch und Portugiesisch hinzukommen. Dafür notwendig ist dann auch eine entsprechende Aufbereitung der Terminologiebestände. Zusätzliche Fachgebiete wie z. B. BWL, Medizin und Recht sollen folgen.

Auf „technischer Seite“ werden in Kürze neue Funktionen wie Dokumenten- und Website-Übersetzungen eingeführt. Langfristig nützlich könnte eine aktive Schnittstelle zum VW-eigenen Terminologiemanagementsystem sein sowie der Aufbau eines eigenen Prozesses „Maschinelle Übersetzung technischer Dokumentation“.

Um die Akzeptanz und Verbreitung maschineller Übersetzung auf der Nutzerseite zu fördern, sind kontinuierliche bewusstseinsbildende Maßnahmen notwendig, um den individuellen Nutzen auf der einen Seite, aber auf der anderen Seite auch die Möglichkeiten und Grenzen maschineller Übersetzung deutlich zu machen („management of expectations“). Des Weiteren müssen die Anwender – auch im Sinne des Unternehmens – für Themen wie Datensicherheit (in Bezug auf maschinelle Übersetzung im Internet) sensibilisiert werden.

Die Corporate-Trust-Studie attestiert der Mehrzahl deutscher Unternehmen einen „erheblichen Nachholbedarf“ in Sachen Problembewusstsein und Sensibilisierung der

eigenen Mitarbeiter in Bezug auf Gefahrenpotenziale und -abwehr. Vielen fehle ein „effektives Informationsschutzkonzept“, insbesondere auch angesichts so genannter „Innentäter“. Als Hauptursachen führt die Studie mangelndes Bewusstsein seitens der Führungskräfte und Mitarbeiter sowie steigendes Outsourcing und Cloud Computing an (2012).

Was nützen also die ausgeklügeltesten Präventivmaßnahmen wie Verschlüsselung von Daten oder E-Mails, zusätzliche Firewalls, Verbot von USB-Sticks und Wechselfestplatten, Funktastaturen und -mäusen, wenn es ganz einfach ausreicht, geheime oder vertrauliche Daten (evtl. sogar zuhause oder auf einer Dienstreise) im Internet zu übersetzen?!

Übrigens wird die Datensicherheitslücke „maschinelle Übersetzung“ weder in der aktuellen Studie von Corporate Trust, noch von VDMA, BSI oder BKA auch nur mit einem Wort erwähnt ...

5 Bibliografie

Austermühl, Frank (2011). „On Clouds and Crowds: Current Developments in Translation Technology“. *T21N – Translation in Transition*. 2011-

09. <http://www.t21n.com/homepage/articles/T21N-2011-09-Austermuehl.pdf>.

Bredow, Rafaela von et al. (2010). „Ende der Privatheit“. *Der Spiegel* 2/2010. 58-69.

Corporate Trust business risk & crisis management (2012). *CYBERWAR. Studie:*

Industriespionage 2012. Aktuelle Risiken für die deutsche Wirtschaft durch Cyberwar. http://corporate-trust.de/pdf/CT-Studie-2012_FINAL.pdf.

DePalma, Donald A. und Nataly Kelly (2009). „The Business Case for Machine Translation.

How Organizations Justify and Adopt Automated Translation“. Common Sense Advisory. Lowell,

Massachusetts. http://www.commonseadvisory.com/Portals/default/Knowledgebase/ArticleImages/090820_R_mt_business_case_Preview.pdf.

Google Terms of Service. Last modified: March 1,

2012. <http://www.google.com/intl/en/policies/terms/>.

Google Übersetzer. <http://translate.google.de/>.

heise.de. „Industriespionage und Produktpiraterie kosten Firmen Milliarden“. 23.04.2012, 15:00 Uhr. <http://www.heise.de/newsticker/meldung/Industriespionage-und-Produktpiraterie-kosten-Firmen-Milliarden-1545751.html>.

Microsoft Bing Translator. <http://www.bing.com/translator/>.

Nogueira, Danilo und Kelli Semolini (2010). „Will We Be Here Tomorrow?“. *Translation Journal*. 14:3. July 2010. <http://translationjournal.net/journal/53mt.htm>.

Porsiel, Jörg (2011). „Maschinelle Übersetzung bei Volkswagen“. *Tagungsband der tekomp-Frühjahrstagung*. Potsdam 14.-15. April 2011. 32-34.

Porsiel, Jörg (2010). „Maschinelle Übersetzung und Datensicherheit“. *MDÜ Fachzeitschrift für Dolmetscher und Übersetzer* 5/2010. 34-36.

Tappel, Tobias und Dorothee Wiegand (2010). „Words don't come easy. Maschinelle Übersetzung auf dem Desktop oder im Web“. *c't* 9/2010, Heft. 120-127.

Thicke, Lori (2011). „Improving MT results. A Study“. *MultiLingual*. January/February 2011. 37-41.

Wikipedia, Die freie Enzyklopädie. Seite „Buchstabensalat“. Bearbeitungsstand: 6. September 2012, 19:49

UTC. <http://de.wikipedia.org/w/index.php?title=Buchstabensalat&oldid=107751179>.

Wooten, Adam (2011). „Can Companies Obtain Free Professional Services through Crowdsourcing?“ *Desert News*. 18. Feb. 2011. <http://www.deseretnews.com/article/print/705366964/Can-companies-obtain-free-professional-services-through-crowdsourcing.html>.

Zetzsche, Jost (2010). „Hostile Takeover? Welcome Addition? Machine Translation Enters the World of the Translator“. *Translation Journal*. 14:3. July 2010. <http://translationjournal.net/journal/53mt1.htm>.

T21N - Translation in Transition

T21N offers a cutting-edge electronic publishing venue, created by experts for both young talent and established researchers from the worlds of translation and interpreting.

T21N provides a stage for emerging ideas and new academic talent to present their ideas in a digital reading site, where speed and ease meet enjoyment.

T21N is exclusively published online at <http://www.t21n.com>.

Articles in compliance with our style sheet may be submitted at any time and will be published at short notice.

T21N editors research and teach at the Institute of Translation and Interpreting at the University of Heidelberg in Germany.

Editors:

Dipl.-Übers. Viktorija Bilić, Dr. Anja Holderbaum,
Dr. Anne Kimmes, Prof. Dr. Joachim Kornelius,
Dr. John Stewart, Dr. Christoph Stoll